N = 1 effect sizes: Comparing models under autocorrelation

R. Manolov and A. Solanas

Department of Behavioral Sciences Methods, Faculty of Psychology, University of Barcelona, Barcelona, Spain, rrumenov13@ub.edu

Generalization of treatment effectiveness from single-case designs is possible whenever findings are replicated across units and integrated by means of meta-analysis. There has already been developed a diversity of measures summarizing the magnitude of the intervention effect: some designed for group studies (e.g., Cohen's *d*), while others specifically destined to N-of-1 designs (e.g., regression-based approaches like Gorsuch's trend analysis [1]; White, Rusch, Kazdin, & Hartmann's *d* [2], Allison & Gorman's model [3], and the non-regression Percent of Nonoverlapping Data; PND). The objective of the current study is to compare the abovementioned techniques in terms of effect detection under the presence of serial dependence and for different design lengths.

Method

The following design lengths were studied: a) N = 10; $n_A = n_B = 5$; b) N = 15; $n_A = 5$; $n_B = 10$; c) N = 15; $n_A = 7$; $n_B = 8$; d) N = 20; $n_A = 5$; $n_B = 15$; e) N = 20, $n_A = n_B = 10$; f) N = 30, $n_A = n_B = 15$.

In order to make possible the simulation of different data patterns (i.e., random fluctuation, level change, slope change, trend, and combination of effects), for each of the aforementioned design lengths data were generated according to the model presented in Huitema & McKean [4]:

$y_t = \beta_0 + \beta_1{*}T_t + \beta_2{*}D_t + \beta_3{*}SC_t + \varepsilon_t$, where:

- y_t: the value of the dependent variable (behavior) at time t;
- β_0: intercept = 0.0;
- $\beta_1, \beta_2, \beta_3$: partial correlation coefficients;
- T_t: value of the time variable at time t (takes values from 1 to N);
- D_t: level change variable (equal to 0 for phase A and equal to 1 for phase B);
- SC_t: value of the slope change variable. $SC_t = [T_t - (n_A + 1)]{*}D_t$. The first n_A data points are equal to zero, while the following ones increment from 0 to ($n_B - 1$).
- ε_t: error term.
- The error term is generated according to: $\varepsilon_t = \varphi_1 {*}\, \varepsilon_{t-1} + u_t$, wherethe autoregressive parameter (φ_1) takes values from −0.9 to 0.9 with a step of 0.1.
- u_t follows a normal distribution with mean zero and unitary standard deviation

The values of the partial correlation coefficients were selected after several trials in such a way as to produce the same mean difference (equal to β_2) between phases for the shortest design ($n_A = n_B = 5$): $\beta_1 = 0.06, \beta_2 = 0.30, \beta_3 = 0.15$. Those values also allowed avoiding floor and ceiling effects.

For each of the experimental conditions defined by the combination of design length, autocorrelation level, and data pattern six effect sizes models were computed 100,000 times and were then averaged across all iterations. The effect size measurements were obtained in terms of R-squared, converting from *d* whenever necessary and using adjusted R-squared for Allison & Gorman's model as suggested by the authors. Solely the PND was not measured in the same scale and comparisons were made on the basis of visual inspection. Fortran 90 programs and NAGfl90 libraries' external subroutines were used for data generation (*nag_rand_seed_set* and *nag_rand_normal*) and multiple regression analysis (*nag_mult_lin_reg*).

Results

Gorsuch's model produced low effect size estimates, ranging from 0.01 to 0.06, concurring with Parker & Brossart [5]. This index produced R-squared that were affected by autocorrelation but did not differentiate between data patterns, and it was the model that showed poorest performance. Varying autocorrelation from −0.9 to 0.9 produced linearily increments in effect size estimates and the most affected techniques where the models of Allison & Gorman and White et al. The influence of serial dependence on PND had a U-shape and was less pronounced than in other techniques.

The regression-based techniques distinguished patterns in a lesser degree and only for long and balanced series. Moreover, the models of Allison & Gorman and White et al. produced seemingly too large R-squared. The two versions of Cohen's *d* performed better and a visual comparison reveals that PND differentiated the most between data patterns for the shortest designs.

Regards design length, as expected, longer data series led to a better differentiation between the effects present in the measurements, although designs such as $n_A = n_B = 5$ and $n_A = 5$, $n_B = 15$ were associated with greater R-squared than $n_A = n_B = 10$ or 15. Consistent with how data were simulated was the fact that changes in slope produced greater R-squared than changes in level.

Discussion

All of the models studied have been criticized on different basis: not being designed for N = 1 designs (Cohen's *d*), being too sensitive to outliers (PND), not taking into account slope change (Gorsuch, White et al.), producing too large R-squared (Allison & Gorman). The results of our simulation study show that simpler methods (like Cohen's *d* and PND) may be more effective than more sophisticated and conceptually more suitable (regression-based) methods. Further research is needed to explore optimal ways of simulating real behavioral data patterns, while other possible lines of future investigations include applying the effect size methods to designs controlling for extraneous variables (e.g., ABAB).

References

1. Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment,* **5**, 141-154.
2. White, D. M., Rusch, F. R., Kazdin, A. E., & Hartmann, D. P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment,* **11**, 281-296.
3. Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy,* **31**, 621-631.

4. Huitema, B. E., & McKean, J. W. (2007). An improved portmanteau test for autocorrelated errors in interrupted time-series regression models. *Behavior Research Methods,* **39**, 343-349.

5. Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy,* **34**, 189-211.